Statistical Inference Course Project Part 1

Alejandro

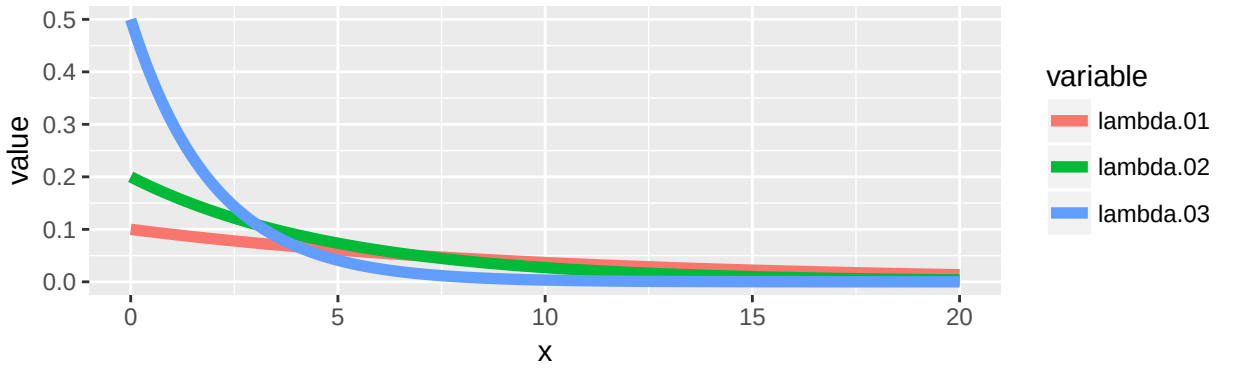
14 Januar 2018

Overview

In this Peer-Graded-Assignment the central limit theorem (CLT) and the ToothGrowth data will be analyzed. For the first part we will rely in the simulation power of R to empirically derive the CLT. In the second part the ToothGrowth data in the R datasets package will be analyzed using techniques learned in class such as confidence intervals and/or hypothesis tests.

Part 1: Simulation Exercise

We will use the exponential distribution and compare the mean of different simulations with the CLT. In the following graph we can see the exponential distribution plotted for different values of lambda.



CTL

To analyze the CTL we will do a set generate 40 random numbers from the exponential distribution and compute the mean and standard deviation for this 40 random numbers. This will be done a total of 1000 times.

```
lambda <- 0.2
n <- 40
sims.mean <- NULL
sims.sd <- NULL
for(i in 1:1000){
    sims <- rexp(n,lambda)
    sims.mean <- c(sims.mean, mean(sims))
    sims.sd <- c(sims.sd , sd(sims))
}
```

In theory the mean and standard deviation for the exponential function should be 1/lambda. With a lambda of 0.2 the mean and standard should be 5 which is really close to what we obtained from the simulations of mean:5.04 and standard deviation: *4.89*

The theoretical standard deviation is defines as (1/lambda)/sqrt(n) which for a lambda of 0.2 is equal to ‘0.7906 and the practical standard deviation is

```
sd(sims.mean)
```

```
## [1] 0.8278644
```

In the following graph we see the the result of the simulations as a histogram as well as the normal distribution. It isnot a perfect fit but with a higher number of simulations the mean of the random numbers will look more and more like the normal distribution.

```
sims.mean <- (sims.mean - mean(sims.mean))/sd(sims.mean)
x <- seq(-4,4, length = 100)
hx <- dnorm(x)
g <- ggplot(data.frame(x = sims.mean), aes(x = x))
g + geom_histogram(aes(y = ..density.. ,fill = I("steelblue")), binwidth = density(sims.mean)$bw) +
    geom_line(data = data.frame(x=x,hx=hx),aes(x=x,y=hx), col = "red", size = 1) +
    labs(x = "Normalized Mean Simulations", y = "Probability", title ="Central Limit Theorem")
```